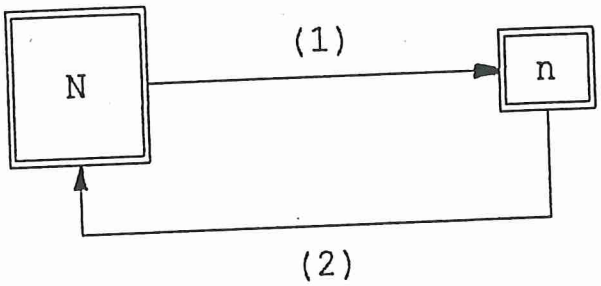
2

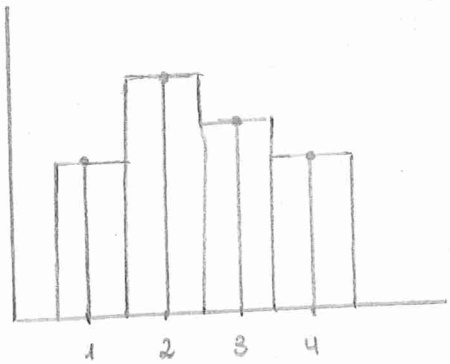
N: población
n: muestra



1.- Teoría del muestreo
2.- Inferencia estadística
3.- Estadística descriptiva:
  · Con la población: PARAMETROS.
  · Con la muestra: ESTADISTICOS.
4.- Teoría de la probabilidad

Estadísticas de TIPO II: discretas.
" " TIPO III: continuas.

¡OJO! * Toda variable discreta se puede transformar en 1 variable continua, considerando q cada 1 de éstas ú discretas es la marca de clase de 1 intérvalo de tamaño unidad



TEMA 1

Tema1: Estadística descriptiva

PRESENTACION Y ORGANIZACION DE DATOS

Una vez recogidos los datos que queremos analizar, debemos ordenarlos de forma que se facilite al máximo su interpretación y/o su posterior análisis.

Para presentar los datos hay que fijarse en:

- el número de observaciones efectuadas
- el número de valores distintos obtenidos. modalidades (si son atributos).

PRESENTACION DE DATOS PARA UNA VARIABLE

* Estadísticas tipo I

- *Pocas observaciones.* → pocos valores.

Ejemplo 1: Edad de cinco personas

Presentación de datos:

EDAD
8
19
23
26
44

* Estadísticas tipo II

Muchos
- *Pocas observaciones*
- *Pocos valores distintos*

Ejemplo 2: Número de personas activas en 50 familias.

Datos observados:
2, 1, 2, 2, 1, 2, 4, 2, 1, 1, 2, 3, 4, 5, 2, 3, 1, ...

Presentación de datos:

Personas activas	Nº de familias
1	16
2	20
3	9
4	5
	50

Concepto de frecuencia: Nº veces que se repite un valor.
presentar datos en 2 columnas
Distintos valores Frecuencia (nº total de observaciones)

* Estadísticas tipo III

- *Muchas observaciones*
- *Muchos valores diferentes*

Ejemplo 3: Importe facturado a cada cliente.

Datos observados:

70325, 880245, 74698, 73652, 64980, 304432, ...

Presentación de datos:

Los datos se presentan AGRUPADOS en INTERVALOS o CLASES. Reducción estadística: Conjunto de datos desordenados → en forma ordenada
Valor representativo ⇒ MARCA DE CLASE

Problemas:

- Número de intervalos. depende del > ó < detalle q. queramos.
- Amplitud de los intervalos:
  - Constante nº de observaciones ≠ en cada intervalo sea la misma en cada intervalo.
  - No constante
- Cómo escribir los límites de los intervalos.
- Error de agrupamiento.

Tema1: Estadística descriptiva

Estadísticas tipo III: ¿ Cómo escribir los límites de los intervalos ?

1ª solución (3a):

Importe de la factura (en unidades monetarias)	Nº de facturas
de 0 a 10	125
de 10 a 20	25
de 20 a 30	11
de 30 a 40	9
de 40 a 50	0
de 50 a 60	0
de 60 a 70	1
de 70 a 80	0
de 80 a 90	0
de 90 a 100	3
de 100 a 110	0
de 110 a 120	0
de 120 a 130	6
de 130 a 140	5
de 140 a 150	3

¡¡ Problema !!: ¿ A qué intervalo pertenece una factura de 10 unidades monetarias ?

Para solucionarlo al pie de la tabla se puede escribir:
De cada intervalo se han excluido las observaciones iguales a su límite superior.

PRECAUCIONES:
- Cálculo de la AMPLITUD de los intervalos
- Cálculo de las MARCAS DE CLASE de los intervalos

Estadísticas tipo III: ¿ Cómo escribir los límites de los intervalos ?

2ª solución (3b):

Importe de la factura (en unidades monetarias)	Nº de facturas
de 0 a 9.999	125
de 10 a 19.999	25
de 20 a 29.999	11
de 30 a 39.999	9
de 40 a 49.999	0
de 50 a 59.999	0
de 60 a 69.999	1
de 70 a 79.999	0
de 80 a 89.999	0
de 90 a 99.999	3
de 100 a 109.999	0
de 110 a 119.999	0
de 120 a 129.999	6
de 130 a 139.999	5
de 140 a 149.999	3

VENTAJAS:
No hay ambigüedad.

INCONVENIENTES:
Se rompe la continuidad de la variable.

PRECAUCIONES:
- Cálculo de la AMPLITUD de los intervalos
- Cálculo de las MARCAS DE CLASE de los intervalos

Estadísticas tipo III: ¿ Cómo escribir los límites de los intervalos ?

3ª solución (3c): emplear LIMITES REALES de clase

Importe de la factura (en unidades monetarias)	Nº de facturas
de -0.5 a 9.5	125
de 9.5 a 19.5	25
de 19.5 a 29.5	11
de 29.5 a 39.5	9
de 39.5 a 49.5	0
de 49.5 a 59.5	0
de 59.5 a 69.5	1
de 69.5 a 79.5	0
de 79.5 a 89.5	0
de 89.5 a 99.5	3
de 99.5 a 109.5	0
de 109.5 a 119.5	0
de 119.5 a 129.5	6
de 129.5 a 139.5	5
de 139.5 a 149.5	3

VENTAJAS:

- No hay ambigüedad.
- No se rompe la continuidad de la variable
- Se obtienen correctamente y con facilidad la AMPLITUD y las MARCAS DE CLASE de los intervalos.

Tema1: Estadística descriptiva

PRESENTACION DE DATOS PARA DOS VARIABLES

De cada 'INDIVIDUO' estudiamos dos caracteres.

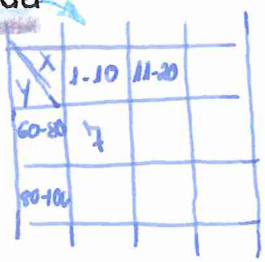
⇩

Para cada 'INDIVIDUO' tendremos un par de valores.

* Estadísticas tipo I : 2 columnas, una para cada variable

* Estadísticas tipo II : se añade la columna de las frecuencias

* Estadísticas tipo III : tabla de doble entrada



PRESENTACION DE DATOS PARA VARIABLES CUALITATIVAS

Las variables cualitativas recogen 'ATRIBUTOS' del individuo estudiado que no se pueden 'MEDIR'.

EJEMPLOS:

Sexo, color de un envase, postura frente al consumo de tabaco, ... se les puedan dar valores ficticiamente: color 1, color 2, color 3...

Los atributos se presentan en MODALIDADES.

Las modalidades han de ser exhaustivas y mutuamente excluyentes.

tienen q. aparecer todas.

q. una observación no pueda pertenecer a + de 1 modalidad

EJEMPLO:

VARIABLE	MODALIDADES
sexo	hombre
	mujer

* Estadísticas de un atributo:

EJEMPLO 4:
Número de productos defectuosos y no defectuosos producidos en el taller A.

modelidades	Nº productos
Defectuosos	10
No defectuosos	1023

* Estadísticas de dos atributos:

- *Tabla de doble entrada*

EJEMPLO 5: (1º atributo ↓ · 2º atributo ↓)
Distribución provincial del número de clientes morosos

modalidades →	Guipúzcoa	Alava	Vizcaya	Navarra	Resto
Morosos	5	3	10	1	10
No morosos	100	50	300	105	197

* Estadísticas de más de dos atributos:

Se usarán tantas tablas de doble entrada como sean necesarias. Ejem. → añadir al ejem. 5 el atributo: el sector productivo al q. pertenece.

Solución ejem.
1. Hacer 1 tabla pa' los clientes del sector construcción.
2. " otra " " " " de cualquier otro sector.

DISTRIBUCION DE FRECUENCIAS

Distribución de frecuencias:

Es una ordenación tabular de los datos en clases con las frecuencias correspondientes.

Frecuencia relativa (f_i %):

Es la frecuencia de clase dividida entre el total de las frecuencias.

Frecuencia acumulada (F_i):

Para estadísticas tipo II: es la frecuencia total de todos los valores menores que ese valor inclusive.

Para estadísticas tipo III: es la frecuencia total de todos los valores menores que el límite superior del intervalo.

Frecuencia relativa acumulada (F_i %):

Es la frecuencia acumulada dividida entre la frecuencia total.

Aplicación al ejemplo 3c:

Importe de la factura	Nº fact. f_i	x_i	F_i	f_i relativa	f_i %	F_i relativa	F_i %
De -0.5 a 9.5	125	4.5	125	0.6649	66.49%	0.6649	66.49%
De 9.5 a 19.5	25	14.5	150	0.1330	13.30%	0.7979	79.79%
De 19.5 a 29.5	11	24.5	161	0.0585	5.85%	0.8564	85.64%
De 29.5 a 39.5	9	34.5	170	0.0479	4.79%	0.9043	90.43%
De 39.5 a 49.5	0	44.5	170	0.0000	0.00%	0.9043	90.43%
De 49.5 a 59.5	0	54.5	170	0.0000	0.00%	0.9043	90.43%
De 59.5 a 69.5	1	64.5	171	0.0053	0.53%	0.9096	90.96%
De 69.5 a 79.5	0	74.5	171	0.0000	0.00%	0.9096	90.96%
De 79.5 a 89.5	0	84.5	171	0.0000	0.00%	0.9096	90.96%
De 89.5 a 99.5	3	94.5	174	0.0160	1.60%	0.9255	92.55%
De 99.5 a 109.5	0	104.5	174	0.0000	0.00%	0.9255	92.55%
De 109.5 a 119.5	0	114.5	174	0.0000	0.00%	0.9255	92.55%
De 119.5 a 129.5	6	124.5	180	0.0319	3.19%	0.9574	95.74%
De 129.5 a 139.5	5	134.5	185	0.0266	2.66%	0.9840	98.40%
De 139.5 a 149.5	3	144.5	188	0.0160	1.60%	1.0000	100.00%
	188						

f_i: frecuencia absoluta. f_i relativa: frecuencia relativa.
x_i: marca de clase. F_i relativa: frecuencia relativa acumulada.
F_i: frecuencia acumulada

Tema1: Estadística descriptiva

REPRESENTACION GRAFICA

Mirar:
1º cuantitativo o cualitativo.
2º si cuantitativo: TIPO I, II ó III

La representación gráfica de los datos tiene como finalidad ofrecer una visión de conjunto del fenómeno que investigamos, más rápidamente perceptible que la lectura directa de los datos numéricos.

REPRESENTACION GRAFICA DE LAS ESTADISTICAS DE CARACTER CUANTITATIVO

REPRESENTACIONES GRAFICAS PARA UNA VARIABLE

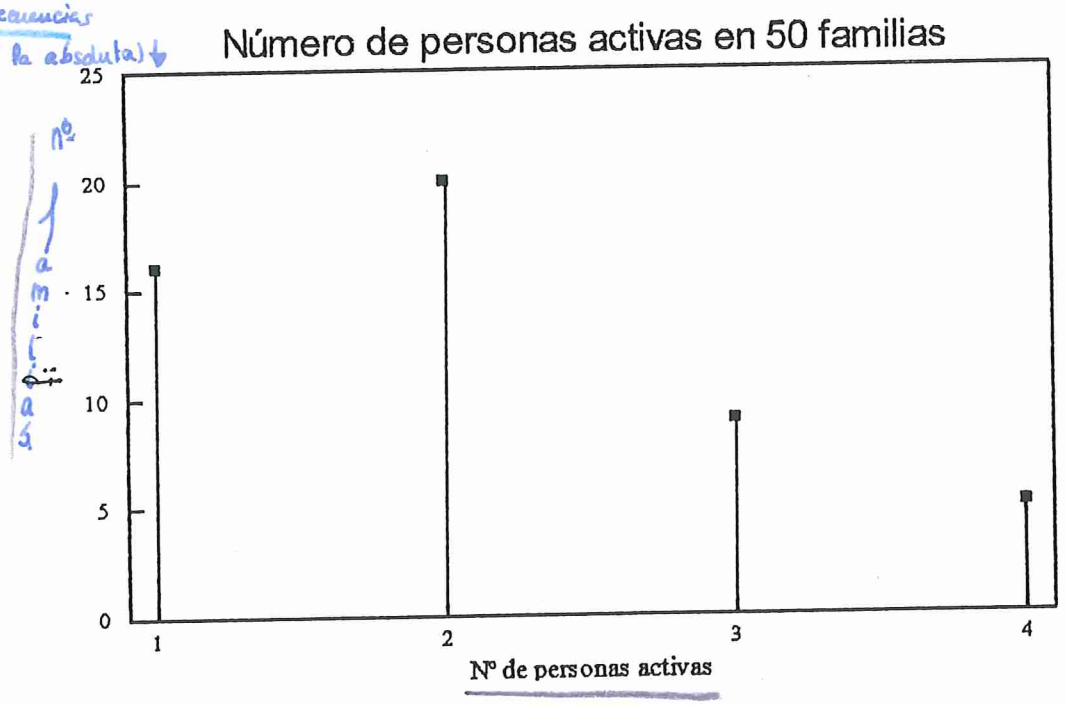
* Estadísticas tipo I:

Su representación gráfica carece de interés.

* Estadísticas tipo II:

Diagrama de barras:
- Ejes cartesianos
- En el eje de abcisas se representan los valores de la variable.
- En el eje de ordenadas se representan las frecuencias.

Aplicación al ejemplo 2:



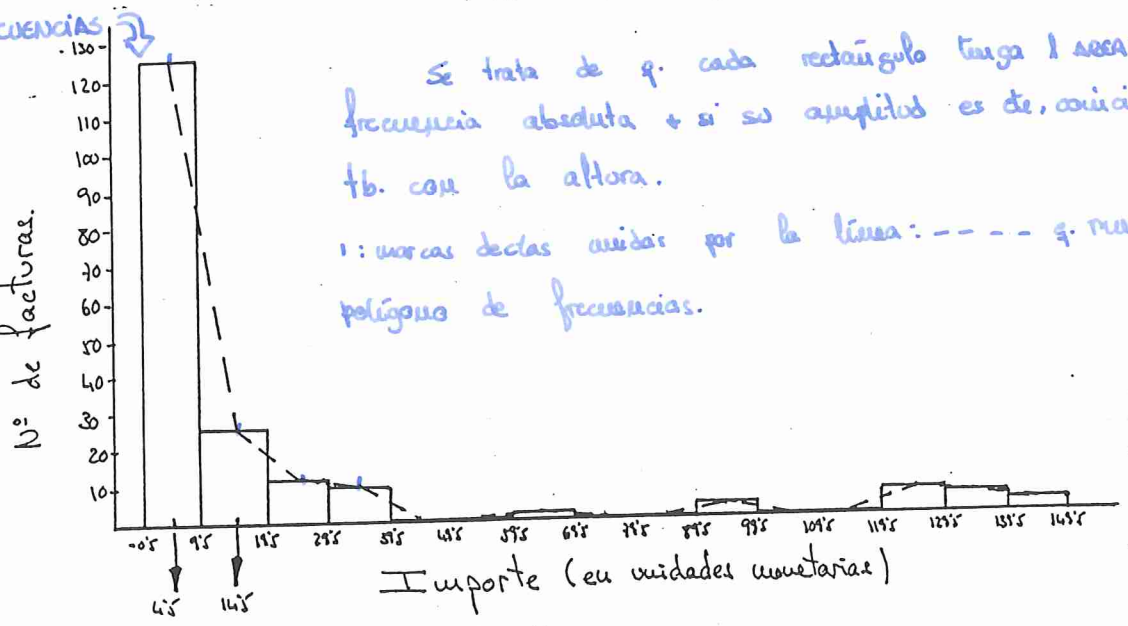
* Estadísticas tipo III:

Representación por áreas.

a) Si los intervalos tienen amplitud constante:
- Ejes cartesianos
- En el eje de abscisas se representan los intervalos de la variable.
- Sobre cada intervalo se levanta un rectángulo de área igual o proporcional a su frecuencia.
- El área total cubierta es igual al número de observaciones.
- El polígono de frecuencias es un gráfico trazado sobre la marca de clase.

Aplicación al ejemplo 3c: Criterio de observación utilizado: OBSERVACIONES POLARES.

- HISTOGRAMA Y POLÍGONO DE FRECUENCIAS



Se trata de q. cada rectángulo tenga 1 AREA = a su frecuencia absoluta + si su amplitud es de, considerar tb. con la altura.

1: marcas dadas unidas por la línea: - - - - q. muestra el polígono de frecuencias.

Haciendo la base igual a uno, el área del rectángulo será igual a la altura; igual a la frecuencia relativa.

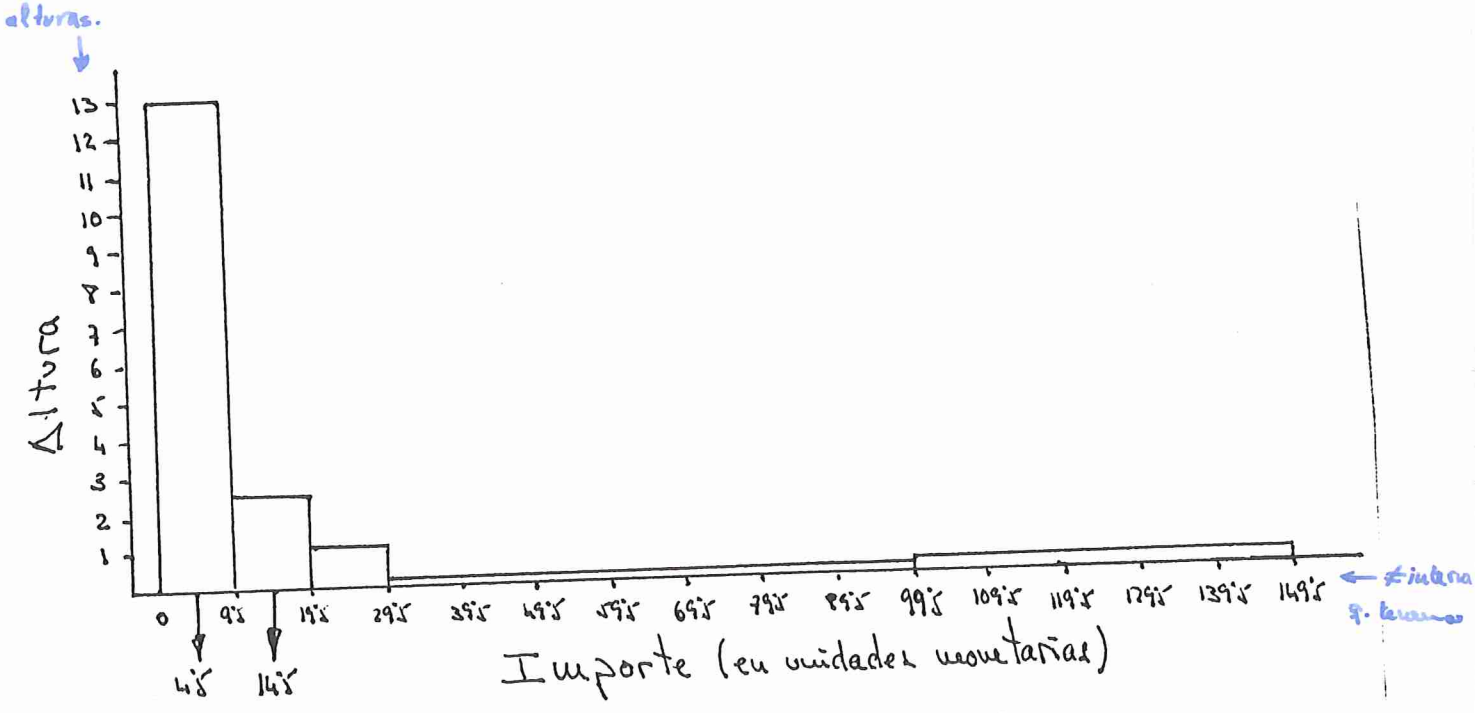
HISTOGRAMA:
Area bajo la línea de trazo continuo.
POLIGONO DE FRECUENCIAS: tiene importancia pues será posible trasformarlo en 1 curva.
Area bajo la línea de trazo discontinuo.

Estadísticas tipo III: su representación gráfica.

b) Si los intervalos tienen amplitud variable:
- Ejes cartesianos.
- En el eje de abscisas se repesentan los intervalos de la variable.
- En el eje de ordendas se miden las 'alturas' de los intervalos. Se calculan dividiendo el área (=frecuencia) entre la base (=amplitud del intervalo).

Ejemplo (3d):

Importe de las fctras.	f_i	x_i	Amplitud	Alturas
De -0.5 a 9.5	125	4.5	10	12.500
De 9.5 a 19.5	25	14.5	10	2.500
De 19.5 a 29.5	11	24.5	10	1.100
De 29.5 a 99.5	13	64.5	70	0.186
De 99.5 a 149.5	14	124.5	50	0.280
	188			



No podemos igualar a 1 pq. no todas las bases son iguales.

REPRESENTACION GRAFICA DE LAS VARIABLES DE CARACTER CUALITATIVO

Los caracteres no se expresan numéricamente. = no son medibles. (si asignamos 1 nº éste será ficticio).
Sólo tenemos modalidades.

DIAGRAMA o GRAFICO DE SECTORES

Representación en un círculo o semicírculo, que se divide en sectores o áreas proporcionales a las intensidades con que se presentan las diversas modalidades del atributo.

Aplicación al ejemplo 4:

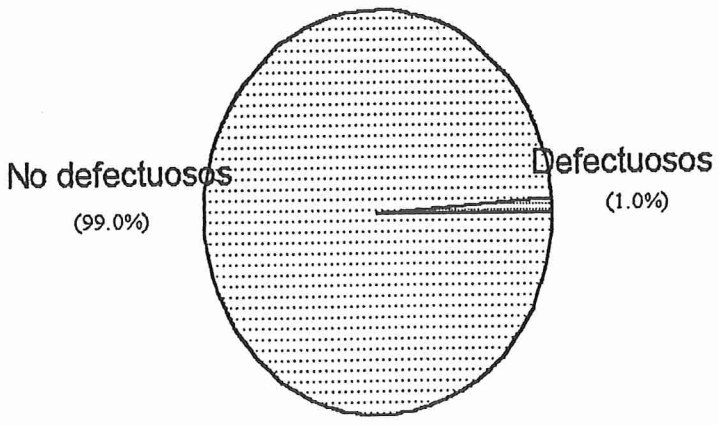
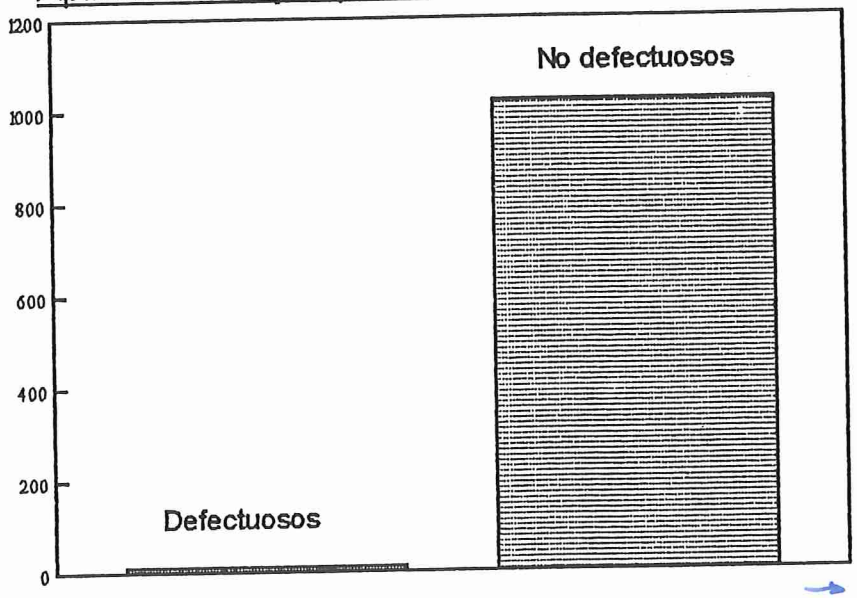


DIAGRAMA DE RECTANGULOS (Gráfica de barras)

- Dos ejes cartesianos.
- En las abcisas se toman las modalidades.
- En las ordenadas se toman las frecuencias.
- Para cada modalidad se toma un segmento de amplitud constante.
- Sobre cada segmento se levantan rectángulos cuyas alturas serán las frecuencias.
- Los rectángulos no se presentan unidos para no dar sensación de continuidad.

Aplicación al ejemplo 4:



→ ≠ modalidades.

Intervalos no unidos pues no existe la continuidad.

Base = 1 ; altura = frecuencia.

REPRESENTACION GRAFICA DE ESTADISTICAS MIXTAS

* No las vamos a utilizar

↳ 2 variables → 1 carácter cuantitativo.
→ 1 " cualitativo.

Se representan tantos histogramas como modalidades tenga el atributo.

DIAGRAMAS ACUMULATIVOS (observaciones con carácter cuantitativo.

Son gráficos de frecuencias acumuladas.

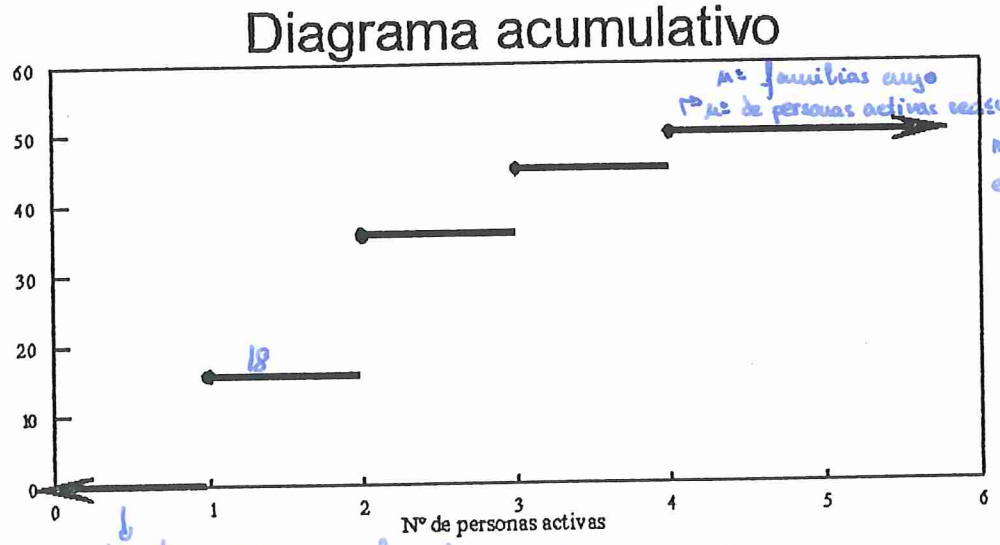
* Estadísticas tipo I

Carecen de interés.

* Estadísticas tipo II

Se representan mediante un diagrama escalonado. q. es 1 diagrama acumulativo.

Aplicación al ejemplo 2:



Dadas 60 familias nº personas activas en cada familia.

b no hay ninguna familia en la q. no haya ninguna persona activa, frecuencia = 0.

* Estadísticas tipo III

Se obtendrá una linea poligonal uniendo los límites superiores de los intervalos en forma ascendente mediante rectas.

Aplicación al ejemplo 3c

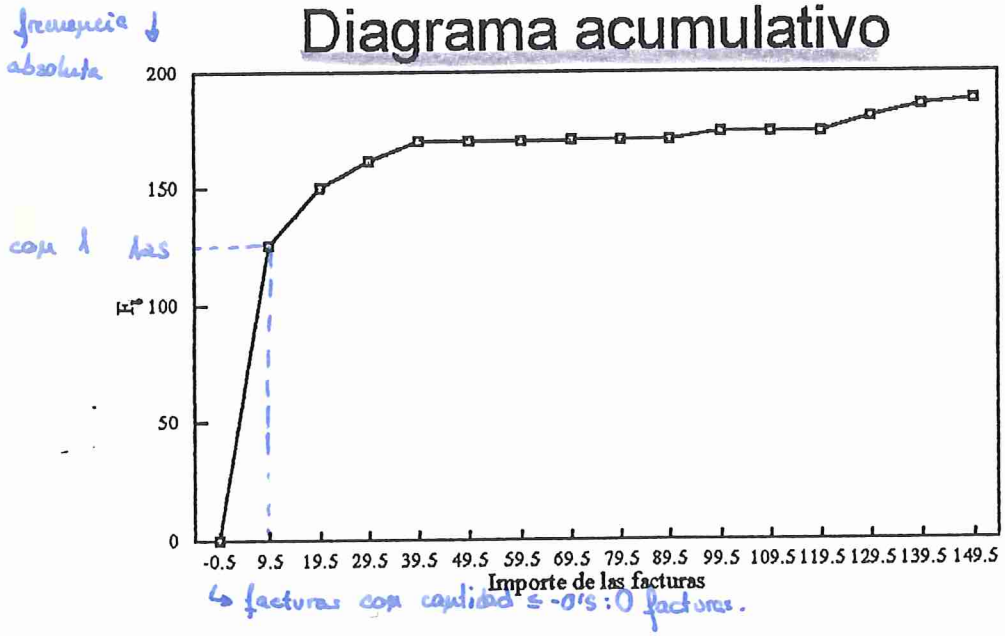


Diagrama acumulativo
frecuencia absoluta
200
150
100
50
0
-0.5 9.5 19.5 29.5 39.5 49.5 59.5 69.5 79.5 89.5 99.5 109.5 119.5 129.5 139.5 149.5
Importe de las facturas
Hay 125 facturas con importe ≤ 9'5.
↳ facturas con cantidad ≤ -0'5 : 0 facturas.

OJIVAS PORCENTUALES

Son diagramas acumulativos en los que se emplean las *frecuencias relativas acumuladas*.

TEMA 2: Continuación del proceso de reducción.

Tema1: Estadística descriptiva Se trata de reducir los datos a una sola cifra.

MEDIDAS DE POSICION

MEDIA ARITMETICA ($\bar{X}$)

* Estadísticas tipo I

$$\bar{X} = \frac{\Sigma x_i}{N}$$

Aplicación al ejemplo 1: Se trata de sumar todas las observaciones y dividir por el nº total de observaciones.

$$\bar{X} = \frac{8+19+23+26+44}{5} = 24$$

→ no está entre esas 5 observaciones, pero tampoco es necesario.

* Estadísticas tipo II

$$\bar{X} = \frac{\Sigma x_i f_i}{N}$$

Donde $N = \Sigma f_i$

sumatorio de los x valores por su frecuencia dividido el nº total de frecuencias.

Aplicación al ejemplo 2:

$$\bar{X} = \frac{1 \cdot 16 + 2 \cdot 20 + 3 \cdot 9 + 4 \cdot 5}{50} = 2.06$$

* Estadísticas tipo III

$$\bar{X} = \frac{\Sigma x_i f_i}{N}$$

Donde: x_i son las marcas *reales* de clase

$N = \Sigma f_i$

Aplicación al ejemplo 3: La misma fórmula q. el TIPO II, pero en vez de utilizar valores se utilizan las marcas de clase.

$$\bar{X} = \frac{4.5 \cdot 125 + 14.5 \cdot 150 + \ldots + 144.5 \cdot 3}{188} = 19.7$$

Importante:

Por este método en realidad se calcula una aproximación a la media, ya que hemos sustituidos los verdaderos valores observados por las marcas de clase de los intervalos. Se comete un error.

MEDIA ARITMETICA PONDERADA

En ocasiones no interesa que todos los valores tengan la misma importancia. Entonces, para calcular la media introduciremos unos "coeficientes de ponderación" que expresen la distinta importancia relativa de cada valor, dando más "peso" a los valores que sean más importantes.

$$\bar{X} = \frac{\Sigma x_i \cdot w_i}{\Sigma w_i}$$

Donde w_i son los coeficientes de ponderación.

MEDIANA (Me)

Es el valor de la variable que ocupa la posición central, una vez ordenadas las observaciones. (bien de > a <, bien de < a >).

* Estadísticas tipo I

Debemos distinguir:

a) si el número de observaciones es impar:

La mediana será el valor de la variable que ocupe la posición central.

b) si el número de observaciones es par:

La mediana será la media aritmética de los dos valores de la variable que ocupen la posición central.

Aplicación al ejemplo 1:

Observaciones ordenadas:

8 19 23 26 44

Número impar de observaciones (5), luego la mediana será el valor de la observación central.

Me = 23

* Estadísticas tipo II

El procedimiento más eficaz para calcular la mediana es a través de las frecuencias acumuladas. La mediana será el valor de la variable que corresponda a la primera frecuencia acumulada mayor que $N/2$

Distinguiremos:

a) Si el número de observaciones es impar (N=2n+1)
La mediana será el valor de la variable correspondiente a la observación número n+1.

Para saber cual es el que q. ocupa el lugar de la mediana N=2n+1; 21=2.13+1 (n+1)

b) Si el número de observaciones es par (N=2n)
La mediana será la media aritmética del valor de la variable correspondiente a la observación número n y el valor de la variable correspondiente a la observación número n+1.

En este caso: N=2n 20=2.10 (n+1)

En ambos casos:
Si $N/2$ coincide con una frecuencia acumulada la mediana será la media arimética del valor de la variable correspondiente a dicha frecuencia acumulada y del siguiente valor.

→ el siguiente valor de $N/2$, una vez ordenados los datos, será ≠ del último valor. (el siguiente valor en ese orden valdrá ≠).

Aplicación al ejemplo 2:

Personas activas	f_i Nº familias	F_i
1	16	16
2	20	36
3	9	45
4	5	50

$N = 50 \rightarrow$ número par

$N/2 = 25$

Valor de la variable correspondiente a la observación número n (= 25): 2

Valor de la variable correspondiente a la observación número n+1 (= 26): 2

$$Me = \frac{2+2}{2} = 2$$

* Estadísticas tipo III

Vamos a ver dos métodos: 1º preocuparnos del intérvalo en el q. está la mediana, para hallarlo utilizamos las frecuencias acumuladas.

A.- Mediante fórmula:

En primer lugar debemos hallar el intervalo crítico que contenga la mediana.

La mediana estará en el primer intervalo cuya frecuencia acumulada sea mayor que $N/2$

A continuación aplicamos la fórmula siguiente:

$$Me = Linf_i + \frac{N/2 - F_{i-1}}{f_i} \cdot h_i$$

Donde:

- el subíndice i nos indica el intervalo crítico
- $Linf_i$ es el límite inferior real del intervalo crítico
- F_{i-1} es la frecuencia acumulada del intervalo anterior al crítico
- f_i es la frecuencia absoluta del intervalo crítico.
- h_i es la amplitud del intervalo crítico

Aplicación al ejemplo 3c: Facturas con ≤ 1 informes.

N: 188

$N/2$: 94

Intervalo crítico: De -0.5 a 9.5

$$Me = -0.5 + \frac{94 - 0}{125} \cdot 10 = 7.02$$

Mediana: Cómo calcularla en las estadísticas de tipo III

B.- Por interpolación lineal → Otra forma de hallar lo anterior.

Para una mejor comprensión es recomendable dibujar el histograma de frecuencias. 1º: localizar el intervalo crítico.

Una vez localizado el intervalo crítico planteamos una regla de tres:

Si a todo el área (= frecuencia) del intervalo crítico le corresponde una base igual a la amplitud del intervalo, al área necesaria para 'alcanzar' la mediana le corresponderá una base 'x'.

Hallado x se suma al límite inferior del intervalo crítico para hallar la mediana.

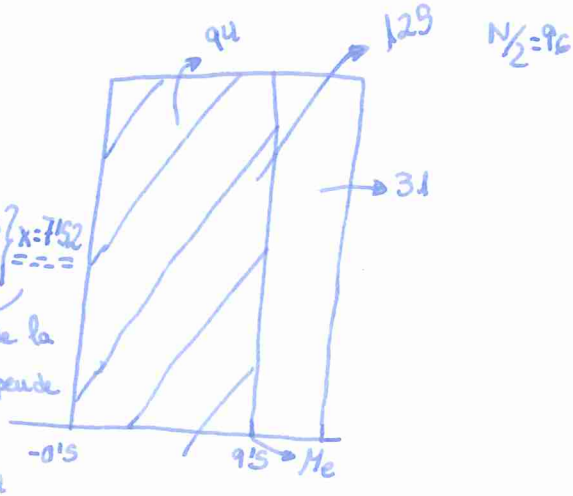
Aplicación al ejemplo 3c:

Intervalo crítico: De -0.5 a 9.5
Area del intervalo: 125
Base del intervalo: 10
La Me se alcanza con un área de
$N/2 - F_{i-1} = 94 - 0 = 94$

Regla de tres:
Si a un área de 125 → una base de 10
a un área de 94 → una base de x

$$x = \frac{94 \cdot 10}{125} = 7.52$$

$$Me = Linf_i + x = -0.5 + 7.52 = 7.02$$



Mediana: Cómo calcularla en las estadísticas de tipo III

Importante:

Por estos métodos en realidad se calcula una aproximación a la mediana, ya que estamos trabajando con datos agrupados.

MODA (Mo)

La moda es el valor de la variable con mayor frecuencia En el ejem. 1 ningún valor se repetía → No hay MODA.
En el ejem. 2. MODA = valor 2.

* Estadísticas tipo I:

La moda es el valor que se repite más veces.

Aplicación al ejemplo 1:

No existe la moda porque no se repite nigún valor.

* Estadísticas tipo II:

La moda es el valor de la variable con mayor frecuencia.

Aplicación al ejemplo 2:

Mo: 2

* Estadísticas tipo III:

En primer lugar, debemos hallar el intervalo modal.

a) Si la amplitud de los intervalos es constante:
El intervalo modal será el de mayor frecuencia

b) Si la amplitud de los intervalos no es constante:
El intervalo modal será el de mayor altura.

Una vez encontrado e intervalo modal podemos buscar una aproximación al valor de la Moda.

Vamos a ver dos métodos:

A.- Tomar como moda la marca de clase del intervalo modal.

B.- Mediante fórmula:

$$Mo = Linf_i + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot h_i$$

Donde:

- $Linf_i$ es el límite inferior real del intervalo modal.
- f_{i-1} es la frecuencia absoluta del intervalo anterior al modal.
- f_{i+1} es la frecuencia absoluta del intervalo posterior al modal.
- h_i es la amplitud del intervalo modal.

Importante:
Por estos métodos en realidad se calcula una aproximación a la moda, ya que estamos trabajando con datos agrupados.

CUANTILES

Son medidas de posición que dividen la distribución en varias partes iguales.

PERCENTILES:
Dividen la distribución en 100 partes iguales.
Son cuantiles de orden 100
Nos dan la situación de los elementos, según el lugar que ocupan en el grupo.

Dos formas para calcularlos:
A.- Mediante fórmula:

si hablamos del percentil 10 será, 10.

$$P_k = Linf_i + \frac{k \cdot N/100 - F_{i-1}}{f_i} \cdot h_i$$

B.- Mediante interpolación: (como la mediana)

DECILES:
Dividen la distribución en 10 partes iguales.
Son cuantiles de orden 10.

CUARTILES:
Dividen la distribución en 4 partes iguales.
Son cuantiles de orden 4. cuartil 1º = percentil 25.

Importante:
Obsérvese que la mediana es un caso particular de los cuantiles.

No tengo fuerza

No tengo fuerza

MEDIDAS DE DISPERSION

DESVIACION MEDIA (DM)

Es la media aritmética de las distancias en valores absolutos de las puntuaciones a la media.
Se utiliza cuando sólo nos interesa conocer la magnitud de las desviaciones.

* Estadísticas tipo I:

$$DM = \frac{\Sigma |x_i - \bar{X}|}{N}$$

* Estadísticas tipo II y III:

$$DM = \frac{\Sigma f_i |x_i - \bar{X}|}{N}$$

En ocasiones, en vez de la media aritmética se empleá la mediana u otros promedios.

es la + utilizada.

VARIANZA (σ^2)

de las desviaciones con respecto a la media.

La varianza es la media aritmética de las diferencias al cuadrado de las puntuaciones con respecto a la media.

* Estadísticas tipo I:

$$\sigma^2 = \frac{\Sigma(x_i - \bar{X})^2}{N}$$

* Estadísticas tipo II y III:

$$\sigma^2 = \frac{\Sigma f_i (x_i - \bar{X})^2}{N}$$

Propiedad importante de la varianza:

* variable aleatoria q. está en f. de otra.

Si $Y = a \cdot x + b$

y denominamos σ_x^2 y σ_y^2 a las varianzas de X e Y respectivamente, se cumple que:

$$\sigma_y^2 = a^2 \cdot \sigma_x^2$$

* Desaparece la "b" pq. es de: dos ctes no tienen dispersión.

DESVIACION TIPICA o estándar (σ)

La desviación típica es la raíz cuadrada de la varianza.

Forma de volver a la medida normal.

$$\sigma = +\sqrt{\sigma^2}$$

COEFICIENTE DE VARIACION (CV)

* Se utiliza para comparar ≠ distribuciones.

El coeficiente de variación expresa la desviación estándar como un tanto por ciento de la media.

Sirve para comparar la dispersión de dos o más distribuciones.

$$CV = \frac{\sigma}{\bar{X}} \cdot 100$$

$\bar{X}$ → ¿media aritmética?

↳ medida relativa de la dispersión con respecto a la media.

Consecuencias si la representamos gráficamente:
- que sea asimétrica. {+ ó −
- que sea simétrica.

MEDIDAS DE ASIMETRIA

+ Si representas? ? tiene forma de campana, es una curva.

Las medidas de asimetría tratan de apreciar la forma global de la distribución.

La asimetría se puede detectar a simple vista, observando el histograma de la distribución.
Además, existen algunos índices que miden la asimetría de la distribución:

A.- Indice de asimetría basado en los cuartiles — indica la simetria - positiva - negativa.

$$A = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2)}$$

⇒

Si A = 0, la distribución puede ser simétrica (el 50% central, pe' no sabe mas nada de los extremos)
Si A > 0, asimetría positiva
Si A < 0, asimetría negativa

B.- Coeficiente de asimetría de Pearson

media - moda media - mediana

$$A_s = \frac{\bar{X} - Mo}{\sigma} = \frac{3\cdot(\bar{X} - Me)}{\sigma}$$

Si $A_s = 0$, no hay asimetría — No podemos asegurar q. la distribución es simétrica. Si podemos asegurar q. si es asimétrica, no lo será mucho.
Si $A_s > 0$, asimetría positiva
Si $A_s < 0$, asimetría negativa

> sesgo a la derecha o a la izq.
sesgo + ó − grande

Podemos decir.

C.- Indice de asimetría basado en los momentos

$$A_3 = \frac{m_3}{\sigma^3}$$

Donde,

$$m_3 = \frac{\Sigma (x_i - \overline{X})^3 \cdot f_i}{N}$$

Si $A_3 = 0$, la distribución puede ser simétrica.
Si $A_3 > 0$, asimetría positiva.
Si $A_3 < 0$, asimetría negativa.

MEDIDAS DE APUNTAMIENTO o CURTOSIS

La curtosis nos indica el mayor o menor achatamiento de la distribución.

Indice de apuntamiento basado en los momentos:

$$A_4 = \frac{m_4}{\sigma^4} - 3$$

Donde,

$$m_4 = \frac{\Sigma (x_i - \overline{X})^4 \cdot f_i}{N}$$

Si $A_4 = 0$, la distribución será *Normal*.
Si $A_4 > 0$, la distribución será *Leptocúrtica*.
Si $A_4 < 0$, la distribución será *Platicúrtica*.

Tema1: Estadística descriptiva

¡¡ SOY LA + MAJA !!

MEDIDAS DE CONCENTRACION

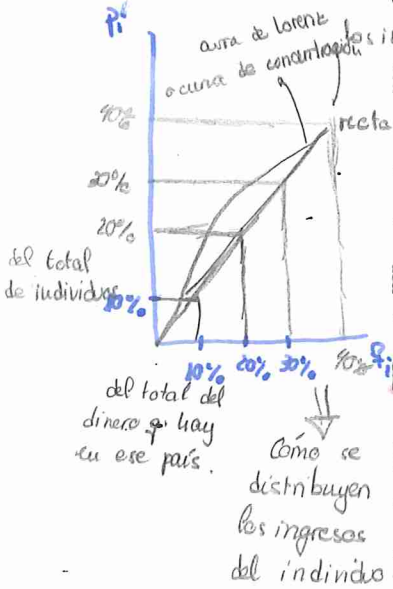
→ se suelen utilizar con variables de tipo económico donde aparecen unidades monetarias.

Mediante las medidas de concentración vamos a estudiar la uniformidad en la distribución de los datos.

Dos métodos:

A.- La curva de concentración o de Lorenz

Es la representación gráfica del fenómeno que estamos estudiando.



los ingresos se reparten por igual entre los individuos: la concentración es mínima.

Debemos añadir unas columnas a la tabla de datos:

· "Volúmenes de" = $x_i \cdot f_i$ (= cantidad total de dinero q. se tiene en cada intervalo).

· q_i: se acumulan los "volúmenes de" y se expresan en porcentaje. (en el eje de abscisas).

· P_i: se acumulan las frecuencias y se expresan en porcentaje. (en el eje de ordenadas).

En un gráfico se representa:

· en abcisas: q_i

· en ordenadas: P_i

La línea poligonal resultante de unir cada par de puntos (q_i;P_i) es la curva de Lorenz.

Recta de equidistribución: es la diagonal del gráfico.

Si la curva de Lorenz se aleja mucho de la recta de equidistribución la concentración será alta. (Habrá poca uniformidad en el reparto de los ingresos de 1 individuo).

la recta no está distribuida uniforme / entre los individuos de la sociedad.

B.- El índice de Gini

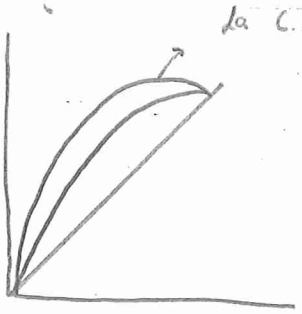
Permite comparar la concentración de dos distribuciones.

$$G = \frac{\sum_{i=1}^{N-1} (P_i - q_i)}{\sum_{i=1}^{N-1} P_i}$$

G oscila entre 0 y 1.

Si G = 0, la concentración es mínima.
Si G = 1, la concentración es máxima. (entre 0 y 1)

Permite comparar por ejem. la concentración entre 2 distribuciones ≠. ¿La renta en la C.A. Vasca está igual / distribuida q. en la C.A. Andaluza?



La C.A. Andaluza tiene > concentración.

* Si las 2 curvas se cortan no se puede decir cuál de las 2 tiene > concentración, por lo q. habrá q. hallar el índice de Gini: